WHAT IS CLAIMED:

1. A method comprising:

specifying a model of a set of biopolymer sequences, the model comprising a first module that characterizes a state of matching between the sequences of the set in a first region and a second module that characterizes a state of matching between the sequences of the set in a second region, wherein the states of matching of the first and second module differ; and

comparing a given set of sequences to the model.

2. The method of claim 1 in which the first module indicates similarity and the second module indicates dissimilarity.

3. The method of claim 1 in which the set consists of two sequences.

4. The method of claim 1 in which the set comprises at least three sequences.

5. The method of claim 1 in which the biopolymer sequences comprise amino acid sequences.

6. The method of claim 1 in which the biopolymer sequences comprise nucleic acid sequences.

7. The method of claim 2 in which the similarity is determined by a similarity scoring matrix.

8. The method of claim 2 in which the dissimilarity is determined by a dissimilarity scoring matrix.

9. The method of claim 8 in which the dissimilarity scoring matrix that is a function of a similarity scoring matrix of the first module.

10. The method of claim 9 in which the dissimilarity scoring matrix is a function of the arithmetic inverse of the similarity scoring matrix.

11. The method of claim 1 in which the model comprises a probabilistic model.

12. The method of claim 11 in which the model expresses a probability that the given set of sequences is a set of sequences of the model.

13. The method of claim 1 in which the model comprises a hidden Markov model.

14. The method of claim 1 in which each of the modules comprises a network of nodes, and each node represents a distribution of monomers at corresponding positions in the sequences of the set.

15. The method of claim 14 in which the distribution of at least some of the nodes of each module differ from each other.

16. The method of claim 14 in which the distribution comprises a function of a scoring matrix that relates occurrences of each monomer at a position in one of the sequences to occurrences of monomers at corresponding positions in at least another one of the sequences.

17. The method of claim 14 in which the network enables positioning of insertions or deletions between a sequence of the set and another sequence of the set.

18. The method of claim 14 in which the network further comprises nodes that represent insertions.

19. The method of claim 7 in which the similarity scoring matrix is a function of independent probabilities of a monomer occurrence.

20. The method of claim 19 in which the distribution $P(a,b)$ of monomers a and b, a scoring matrix $S(a,b)$, and independent probabilities of monomers, $Q(a)$ and $Q(b)$ are related such that $S(a,b) = \log(P(a,b) / (Q(a)\ Q(b))$.

21. The method of claim 1 in which the model further comprises a third module that characterizes the state of matching between each sequences of the set in a third region.

22. The method of claim 21 in which the model comprises a third module that indicates the similarity between a third region of each sequence of the set and a sequence profile.

23. The method of claim 22 in which the sequence profile is indicated by altered scoring matrices.

24. The method of claim 23 in which the sequence profile comprises a profile of a modification site.

25. The method of claim 23 in which the sequence profiles comprises a profile of a processing site.

26. The method of claim 21 in which the third module is positioned between the first and second module with respect to the order of the sequences.

27. The method of claim 25 in which the processing site indicates a preference for at least a basic residue.

28. The method of claim 27 in which the processing site indicates a preference for at two basic residues.

29. The method of claim 25 in which the processing site comprises a convertase processing site.

30. The method of claim 25 in which the processing site comprises a secretase processing site.

31. The method of claim 21 in which the third module is trained.

32. The method of claim 1 in which the sequences of the given set comprises sequences from different species.

33. The method of claim 32 in which the different species comprise mammalian species.

34. The method of claim 31 in which the third module is trained to encompass a family of conserved sequence segments.

35. The method of claim 6 in which the sequences comprise genomic nucleic acid sequences.

36. The method of claim 6 in which the sequences comprise non-coding regions.

37. The method of claim 6 in which the sequences comprise regulatory regions.

38. The method of claim 6 in which the sequences comprise transcriptional regulatory regions.

39. A medium carrying a model capable of enabling a machine to perform comparisons of a set of biopolymer sequences to the model, the model comprising a first module that characterizes a state of matching between the sequences of the set in a first region and a second module that characterizes a state of matching between the sequences of the set in a second region, wherein the states of matching of the first and second module differ.

40. The medium of claim 39 in which the first module characterizes a state of dissimilarity between the sequences of the set.

41. A method comprising:

defining a sequential pattern of biopolymer sequence segments, the pattern comprising a similar segment and a dissimilar segment;

comparing a first biopolymer sequence to a reference to identify similar and dissimilar segments in the first sequence; and

determining if the similar and dissimilar segments of the first biopolymer sequence match the defined sequential pattern.

42. The method of claim 41 in which the comparing and the determining are concurrent.

43. The method of claim 41 in which the reference comprises a second biopolymer sequence.

44. The method of claim 41 in which the reference comprises a sequence profile.

45. The method of claim 41 further comprising repeating the comparing and determining for a plurality of sequences.

46. The method of claim 41 further comprising repeating the comparing and determining such that multiple combinations of sequences selected from a plurality of sequences are compared.

47. The method of claim 46 in which the plurality of sequences comprises sequences from different species of the same phyla.

48. The method of claim 47 in which the plurality of sequences comprises sequences from different mammalian species.

49. The method of claim 47 in which each of the multiple combinations includes sequences from different species.

50. The method of claim 41 in which the determining comprises identifying a value that evaluates the matching to the defined sequential pattern.

51. The method of claim 46 further comprising ranking the combinations based on the identified value.

52. The method of claim 41 further comprising, if the similar and dissimilar segments of the first biopolymer sequence match the defined sequential pattern, assaying a biopolymer that comprises one of the segments of the first biopolymer sequence for an activity.

53. The method of claim 52 in which the biopolymer comprises the similar segment.

54. The method of claim 52 in which the biopolymer comprises the first polymer sequence.

55. A method comprising:

evaluating sets, each set comprising a first sequence from sequences of a first species and a second sequence from sequences of a second species, the evaluating comprising

(i) comparing the first and second sequence of each set to identify similar and dissimilar segments; and

(ii) returning a value indicative of the match between the similar and dissimilar segments of the set and a defined pattern of similarity and dissimilarity; and

identifying sets which return values that exceed a threshold.

56. The method of claim 55 in which the first species is a eukaryotic species.

57. The method of claim 56 in which the first species is a vertebrate species.

58. The method of claim 57 in which the first species is a mammalian species.

59. The method of claim 58 in which the first species is a human.

60. The method of claim 58 in which the second species is a mammalian species.

61. The method of claim 55 in which the similar segment is between processing sites.

62. The method of claim 55 in which the similar segment is adjacent to a processing site.

63. The method of claim 61 in which the dissimilar segment is outside the processing sites.

64. The method of claim 62 in which the processing site is a protease cleavage site.

65. A method comprising:

a) comparing a query sequence to each candidate sequence of a plurality of candidate sequences by a method comprising

i) identifying a first segment in the candidate sequence and a first segment in a query sequence;

ii) determining a first measure that is a measure of the similarity between the first segments; and

iii) determining a second measure that is a measure of the similarity between segments of the query sequence and the candidate sequence, the segments being other than the first segment; and

b) identifying a selected candidate sequence from the plurality of candidate sequences, wherein a comparison of the first and second measures of the selected candidate sequence indicate at least a threshold value.

66. The method of claim 65 in which each first segment is adjacent to a processing site.

67. The method of claim 66 in which the processing site is a convertase processing site.

68. The method of claim 65 in which the first segment is between a first processing site and second site that is a second processing site, a signal sequence, or a carboxy terminus.

69. The method of claim 65 in which the identifying comprises aligning the query sequence and the candidate sequence.

70. The method of claim 69 in which the aligning comprises maximizing local alignments.

71. An article of machine-readable media having encoded thereon software configured to cause a processor to:

a) compare a query sequence to each candidate sequence of a plurality of candidate sequences by a method comprising

i) identifying a first segment in the candidate sequence and a first segment in a query sequence;

ii) determining a first measure that is a measure of the similarity between the first segments; and

iii) determining a second measure that is a measure of the similarity between segments of the query sequence and the candidate sequence, the segments being other than the first segment; and

b) identify a selected candidate sequence from the plurality of candidate sequences, wherein a comparison of the first and second measures of the selected candidate sequence indicate at least a threshold extent of localized similarity.